EXHIBIT 24

BACKGROUND MOSAICKING FOR LOW BIT RATE VIDEO CODING

Frédéric Dufaux[†] *and Fabrice Moscheni*[‡]

[†]Digital Equipment Corporation, Cambridge Research Lab
Cambridge, MA 02139, USA

[‡]Signal Processing Laboratory, Swiss Federal Institute of Technology
CH-1015 Lausanne, Switzerland

ABSTRACT

This paper proposes a new technique to build a background memory based on mosaicking. More precisely, the technique first identifies background and foreground regions based on local motion estimates. Camera motion is then estimated on the background by applying a parametric global motion estimation. Finally, after compensating for camera motion, the background content is temporally integrated in long-term memory. The method leads to high coding performances and allows for content-based functionalities.

1. INTRODUCTION

Efficient integration and representation of the motion information is a key component in a video coding scheme. For this purpose, this paper proposes a new technique to build a background memory based on mosaicking. This method leads to improved coding performances and supports content-based functionalities. Furthermore, it is broadly applicable both in classical motion compensated block-DCT video coding schemes as well as in object-based video coding schemes.

One of the most pertinent problem in video coding lies in the coding of uncovered background areas. In this case, the classical motion compensated prediction scheme is unable to predict newly appearing areas and hence leads to poor performances. To overcome this problem, background memory techniques have been proposed [1, 2]. More precisely, these techniques identify still regions as background and stored them in a long term memory. Whenever an area is uncovered, and providing that it has been visible in the past, the information unavailable with classical prediction techniques can be retrieved from the background memory. These algorithms are effective for video-conference and video-phone sequences which are characterized by a still background. However, the model of a still background does no longer hold for more complex scenes which include camera motion (e.g. pan or zoom).

In order to integrate temporal information, mosaic representations [3, 4], also referred to as salient still [5], have shown to be efficient. Basically, these techniques estimate the camera motion through global motion estimation and align the images in the sequence by canceling the contribution of camera motion. Then, the mosaic is built by temporal integration of the aligned images. In this way, the mosaic capture the information in multiple frames of a video sequence. However, these methods are applied without distinction between background and foreground, though the camera motion is representative of the background motion only. Therefore, foreground objects appear blurred out. Furthermore, as foreground objects are included in the mosaic representation, the problem of uncovered background remains unsolved.

Taking into account the above considerations, this paper combines the ideas of background memory and mosaic representation. More precisely, it proposes a new technique to dynamically build a mosaic of the background, where the latter is now defined as the regions whose motion is coherent with the camera motion. Straightforwardly, as the proposed technique compensates for camera motion, it overcomes the shortcoming of the classical background memory techniques. Furthermore, as the temporal integration is performed on the background region only, the technique outperforms classical mosaic representations.

This technique requires three main stages. A discrimination is first made between background and foreground regions based on motion information. Second, the camera motion is estimated on the background by global motion estimation. Third, using the camera motion information, a dynamic mosaic of the background is progressively integrated and stored in a long term memory. Each of these stages represents a challenging problem, and this paper provides also efficient solutions to solve them.

Using the background mosaicking technique for video coding allows for an efficient integration and representation of the motion information. More specifically, a two-way motion compensated prediction is introduced. The current frame can be predicted either from the background memory using global motion estimates or from the previous frame using local motion estimates. Therefore, this technique handles to a large extent the problem of uncovered background. Furthermore, the amount of side information required to represent the motion information is greatly reduced due to the compact parametric representation of camera motion. Those gains are obtained without introducing any coding delay, though an additional frame buffer is required. The method can be applied independently from the subsequent coding technique. In particular, though the method is applicable in classical motion compensated block-DCT coding schemes, it is also very appropriate for an object-based representation of the scene. Finally, the method provides a high resolution panoramic view of the background.

2. BACKGROUND MOSAICKING

As illustrated in Fig. 1, the proposed background mosaicking technique is composed of the following steps: background/foreground segmentation, global motion estimation on the background region, and dynamic mosaic representation of the background. These steps are now described in more details.

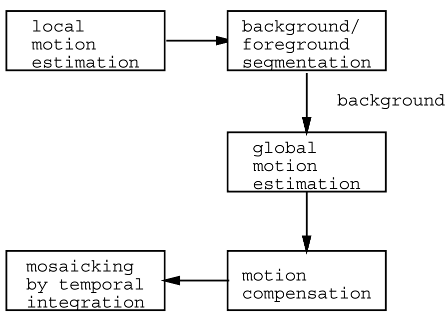


Figure 1: Background mosaicking.

2.1. Segmentation of background and foreground

The discrimination between background and foreground is based on local motion information. A local motion vector field is first estimated. A dominant motion is extracted by a clustering of the motion vectors. Then, regions moving according to the dominant motion are identified as background, and otherwise as foreground.

Most naturally, the discrimination is based on the local motion vector estimates. If a motion vector is close to the dominant motion, the pixel locations characterized by this motion vector are assigned to the background. However, in low contrast areas, the local motion estimation is underconstrained and the resulting motion vectors are unreliable. In this case, the above classification based on motion similarity may fail. To overcome this difficulty, the residual information present in the prediction error is also exploited [6]. Namely, if, for an area, the residual obtained with the dominant motion is similar to the one obtained when using the local motion vector, then this area is also assigned to the background. In this way, low contrast areas where motion estimation is unreliable can be robustly classified as background or foreground. The background/foreground segmentation algorithm is summarized in Fig. 2.

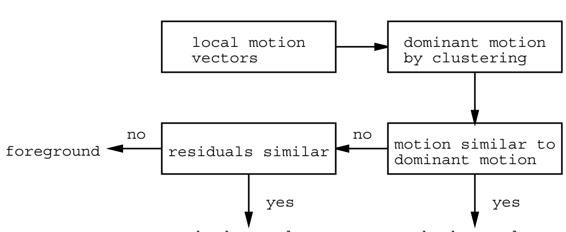


Figure 2: Background/foreground segmentation.

2.2. Camera motion estimation

After discriminating between background and foreground regions, camera motion is robustly estimated on the background. In this way, the camera motion estimate is not spoilt by the presence of outliers due to foreground objects whose motion is not representative of the camera motion.

The camera motion is modeled by a parametric motion model. For this purpose, two models are widely used, the affine model and the perspective model [7]. The affine model is expressed as

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} a_1 + a_2 x + a_3 y \\ a_4 + a_5 x + a_6 y \end{pmatrix} , \qquad (1)$$

and the perspective model is given by

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \frac{a_1 + a_2 x + a_3 y}{a_7 x + a_8 y + 1} \\ \frac{a_4 + a_5 x + a_6 y}{a_7 x + a_8 y + 1} \end{pmatrix} . \qquad (2)$$

where $(x, y)^T$ is the pixel location in the previous frame and $(x', y')^T$ is the corresponding location in the current frame.

The 6-parameters affine model allows for the representation of the motion of a planar surface under orthographic projection. However, it makes the assumption that the depth map of the scene is small relative to the distance to the camera. In order to relax this constraint, the 8-parameters perspective model is rather used in this paper. This model allows for the representation of the motion of a planar surface under perspective projection.

In order to robustly estimate the motion parameters, a matching technique is used [7]. This technique has been shown to outperforms differential and linear regression techniques [8]. The motion parameters $\vec{a} = (a_1, \ldots, a_n)$ are obtained by minimizing a disparity measure between the region R in the current frame and the mapped region in the previous frame

$$\min_{\vec{a}} \sum_{\vec{r} \in R} \parallel I(\vec{r}, t) - I(T(\vec{r}, \vec{a}), t-1) \parallel , \qquad (3)$$

where $I(\vec{r}, t)$ denotes the image intensity at location $\vec{r}$ and time t, $T(\vec{r}, \vec{a})$ is the location to be matched in the previous frame.

The search is carried out in the n-dimensional parameters space. To decrease computational complexity, a fast non-exhaustive search is carried out and the motion parameters are estimated progressively. First the translational component is computed (i.e. a_1 and a_4), then the affine parameters (i.e. a_2, a_3, a_5 and a_6), and finally the perspective components (i.e. a_7 and a_8). Besides decreasing complexity, this search strategy tends to produce more robust estimates, as the translational component carries usually more information than the other parameters which reduce sometimes to little more than noise. However, this search strategy obviously induces the risk to be trapped in a local minimum.

Furthermore, the matching motion estimation algorithm is applied on a multiresolution structure based on a Gaussian pyramid. The final motion parameters at one level propagate as initial estimates on the next level. This multiresolution scheme allows for the reduction of the computational load, as well as the prevention of local minima due to the non-exhaustive search.

2.3. Dynamic mosaic of the background

Once the background has been segmented and the camera motion estimated, the background information is temporally integrated by mosaicking. In [3], a distinction is made between static and dynamic mosaics. The static mosaic integrates the information of all the frames in a video segment (e.g. mean, median, weighted mean, weighted median). Conversely, the dynamic mosaic corresponds to a progressive update of the mosaic content by gradually integrating the information of the individual frames. As the static mosaic puts much more constraints on the video coding scheme (high coding delay, buffering of many frames), the dynamic mosaic is rather considered in this paper.

To build the mosaic, the frames are aligned with respect to a coordinate system. This coordinate system corresponds either to a fixed reference (e.g. the first frame or a prefered view), or to a time-varying reference (e.g. the current frame).

In the case of a fixed reference, the background region of the individual frames are dynamically integrated in the mosaic using the following update strategy

$$M(\vec{r},t) = (1-\alpha\delta(\vec{r}))M(\vec{r},t-1) + \alpha\delta(\vec{r})I(T(\vec{r},\Sigma a(t)),t) \ , \quad (4)$$

where $M(\vec{r},t)$ denotes the dynamic mosaic at location $\vec{r}$ and time t, $\Sigma_t a(t)$ is the sum of the global motion parameters obtain pairwise between consecutive frames over the time period between the reference and the current frames, α is a weighting factor such as $\alpha \in [0,1]$, and $\delta(\vec{r}) = 1$ if $\vec{r}$ belongs to the background and 0 otherwise.

Very similarly, in the case of a time-varying reference corresponding to the current frame, the mosaic update becomes

$$M(\vec{r},t) = (1-\alpha\delta(\vec{r}))M(T(\vec{r},\vec{a}),t-1) + \alpha\delta(\vec{r})I(\vec{r},t) \ . \quad (5)$$

The parameter α controls the update of the mosaic, it can be adapted on the fly. For instance, if a measure of the reliability of the global motion estimation is available (in our case the matching error), α can be decreased (respectively increased) when the estimate is unreliable (respectively reliable).

When the sole goal of the background mosaicking technique is to achieve better motion compensated prediction, Eq. (5) leads to higher performances as it remains closer to the frame to be predicted. However, when a panoramic view of the background from a specific view point is desired, Eq. (4) has to be prefered.

2.4. Application to video coding

In the framework of coding, the background mosaic allows for a two-way motion compensated prediction as illustrated in Fig. 3. For instance, in a motion compensated block-DCT scheme, each block of the background can be predicted either from the background mosaic (using global motion) or from the previous frame as in the classical scheme (using local motion), depending on which predictor leads to the lower prediction error. This way, uncovered background areas can be retrieved from the background mosaic providing that they have been visible in the past.

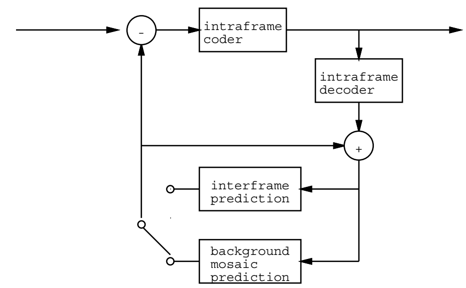


Figure 3: Two-way prediction.

Furthermore, the parametric global motion estimation efficiently handles camera motion such as panning and zooming, resulting in a reduced motion side information. Finally, the process of integrating many frames in the mosaic is filtering the noise in the sequence, leading to better prediction.

Besides, the background mosaic may also be used to construct an artificial high resolution panoramic view of the background similarly to [9].

3. EXPERIMENTAL RESULTS

Experimental results are presented in this section. Two sequences in CIF format have been used. Fig. 4 shows a frame of the sequence "Stefan" which is characterized by large camera motion (pan and zoom), the dynamic background mosaic after integrating 200 frames, and the gain obtained on the prediction when using background mosaicking and two-way motion compensated prediction. Fig. 5 shows similar results for the sequence "Weather" which is composed of a still background and a moving foreground object. In the above results, the mosaic has been build using Eq. (5). As far as the interframe prediction is concerned, a classical block matching motion estimation has been used, with a block size of 8×8 pixels, a maximum displacement of ± 25 pixels, and half-pel accuracy.

It can be observed that the method generates a panoramic view of the background where foreground objects have partly disappeared. In terms of prediction, the proposed method leads to an average gain of approximately 1dB on both sequences, with peaks up to 2dB and 4dB.

4. CONCLUSION

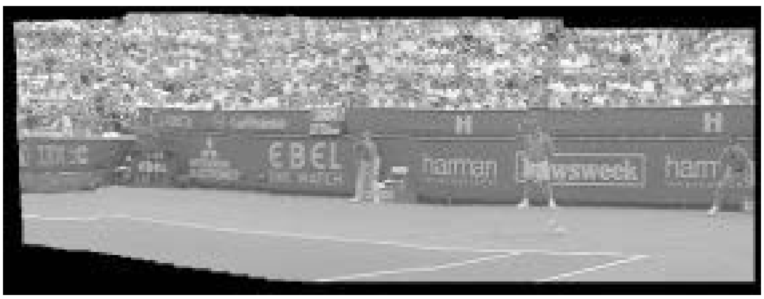
This paper described a new technique to incrementally integrate frames in a background mosaic. Simulation results have shown that the method leads to significantly better performances in a motion compensated video coding scheme. The technique provides also a panoramic view of the background.

5. REFERENCES

[1] D. Hepper. Efficiency analysis and application of uncovered background prediction in a low bit rate image coder. *IEEE Trans. Commun.*, vol. COM-38, no. 9, pp. 1578-1584, September 1990.

a)

b)



c)

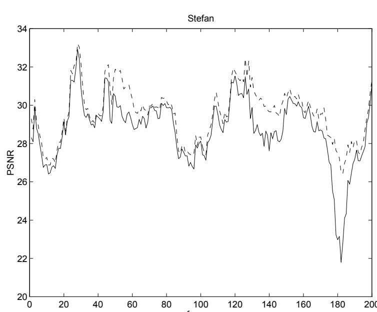


Figure 4: "Stefan": a) a frame, b) background mosaic after 200 frames, c) motion compensated prediction (continuous line: classical interframe prediction, dashed line: two-way mosaic prediction).

a)

b)

c)

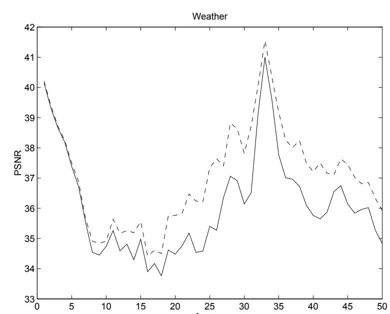


Figure 5: "Weather": a) a frame, b) background mosaic after 50 frames, c) motion compensated prediction (continuous line: classical interframe prediction, dashed line: two-way mosaic prediction).

[2] X. Yuan. Hierarchical uncovered background prediction in a low bit-rate video coder. In *Picture Coding Symposium '93*, page 12.1, Lausanne, Switzerland, March 1993.

[3] M. Irani, S. Hsu, and P. Anandan. Mosaic-based video compression. In *SPIE Proc. Digital Video Compression: Algorithms and Technologies*, volume 2419, San Jose, CA, February 1995.

[4] H.S. Sawhney, S. Ayer, and M. Gorkani. Model-based 2D & 3D dominant motion estimation for mosaicing and video representation. In *Int. Conf. on Computer Vision*, pages 583–590, Cambridge, MA, June 1995.

[5] L.A. Teodosio and W. Bender. Salient video stills: content and context preserved. In *Proc. ACM Int. Conf. on Multimedia*, Anaheim, CA, August 1993.

[6] F. Moscheni and F. Dufaux. Region merging based on robust statistical testing. In *SPIE Proc. Visual Communications and Image Processing'96*, Orlando, March 1996.

[7] F. Dufaux and F. Moscheni. Segmentation-based motion estimation for second generation video coding techniques. In L. Torres and M. Kunt, editors, *Video Coding: A Second Generation Approach*, pages 219–263. Kluwer Academic Publishers, 1996.

[8] F. Moscheni, F. Dufaux, and M. Kunt. A new two-stage global/local motion estimation based on a background/foreground segmentation. In *IEEE Proc. ICASSP'95*, Detroit, MI, May 1995.

[9] R. Kermode and A. Lippman. Coding for content: enhanced resolution from coding. In *IEEE Proc. ICIP'95*, Washington, DC, October 1995.